Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnline CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 55-4 (2022) 19–24

System identification and mathematical modeling of the pandemic spread COVID-19 in Serbia

Jasmina Lozanović Šajić,*, Sonja Langthaler,* Jörg Schröttner* and Christian Baumgartner***

**Institute of Health Care Engineering with European Testing Center of Medical Devices, Graz University of Technology, Austria (e-mail: j.lozanovicsajic@tugraz.at)*

*** Innovation Center of the Faculty of Mechanical Engineering, University of Belgrade, Kraljice Marije 16, Belgrade, Serbia (e-mail: jlozanovic@mas.bg.ac.rs)*

Abstract: This paper presents applications of control system theory in biomedical engineering. These methodologies are used in engineering sciences to obtain a mathematical model of systems, but system identification as scientific methodology is rarely used in biomedical engineering. The paper presents exemplarily control theory and system identification as methods for obtaining a mathematical model of the spread SARS-CoV-2 virus. The models obtained in the course of this are data-driven and strongly data-dependent. The available dataset allowed us to consider a model of a pandemic spread in the context of both the number of tested individuals and the number of infected individuals and with a resultant model that is nonlinear. We also considered a mathematical model for the dependence between the number of confirmed infected individuals and the number of deaths caused by the disease. The resulting model is linear given with the transfer function corresponding to the second-order differential equation. The mathematical models developed were additionally analyzed in accordance with controllability and observability.

Keywords: biomedical engineering, system identification, mathematical model, covid-19

1. INTRODUCTION

This paper presents a novel approach of the application of control system theory and system identification in biomedical engineering and medical sciences. Control system theory is commonly used today in advanced medical devices. This research presents possibilities of application system identification methodology to develop mathematical models for different biomedical systems types such as the spread of a virus. The models presented here are dynamic, and physical and mathematical laws can be used to develop these models whenever possible. However, if it is impossible to develop a mathematical model using known formulas, we can use the methodology from control system theory – system identification. In this paper, the control theory and system identification knowledge was applied to develop a spread model for SARS-CoV-2.

This research considers the mathematical model of the spread SARS-CoV-2. In December 2019, China reported a cluster of cases of pneumonia in Wuhan, Hubei Province. The responsible pathogen is a novel coronavirus named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The first case in Europe of confirmed infection, coronavirus disease 2019 (COVID-19), was diagnosed on January 24th, 2020. This novel coronavirus and disease were unknown before the outbreak began in Wuhan, China, in December 2019, World Health Organization.

The system identification is based on observed and/or measured system data, Ljung (2002). The system is defined with input and output vectors, but also, in exceptional cases, control systems can be single input and single output systems, also called SISO systems. Physical systems are mostly nonlinear, Dorf et al. (2010). The linearization of some nonlinear models in engineering and biology thus satisfies the criteria for analysis. For example, a heart model based on the cardiac output curve is usually nonlinear, linearisation simplified this system of cardiac output regulation, Khoo (2018).

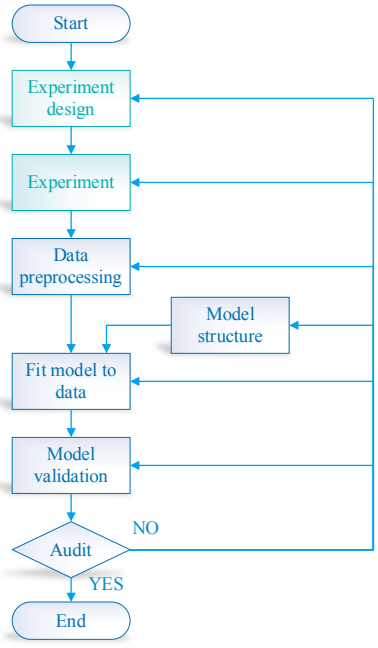

Figure 1. System identification cycle

10.1016/j.ifacol.2022.06.003

System identification aims are to develop a mathematical system model from measured input and output data, Schoukens et al. (2012). The key elements in the system identification cycle are experiment design, experiment, data pre-processing, model structure choice, fit model to data, and model validation. A general algorithm of system identification with the key elements is shown in Figure 1, Verhaegen et al. (2007). In addition, continuous-time differential equations can describe physical models in general.

Nowadays, system identification has been simplified by using state-of-the-art modern software tools in MATLAB. In general, system identification in MATLAB includes system identification approach, model structures and properties, time and frequency domain measured data, importing and processing data, representing data, filtering data, black box modeling linear and nonlinear models, identifying process models, etc., Ljung et al. (2014).

2. SYSTEM MODELING

This paper presents a system identification application for bioengineering and medical purposes. All models under consideration are based on available data. In the first example, we consider developing a mathematical model of the spread caused by SARS-CoV-2. Most epidemiological models are based on the rate of changes in susceptible, infected, and recovered individuals, also known as SIR models. Classical epidemiological models are fundamentals for understanding, predicting, and planning control strategies for an epidemic, Giordano et al. (2020).

The SIR model is given by Eq. (1)

$$S(t) + I(t) + R(t) = N \tag{1}$$

where S(t) is the number of susceptible individuals, I(t) is the number of infected individuals, R(t) is the number of recovered individuals, and N is the population.

Therefore, the SIR model is described with the rate of changes S(t), I(t), and R(t), Weiss et al. (2013),

$$\begin{aligned} \frac{dS}{dt} &= -\beta IS \\ \frac{dI}{dt} &= \beta IS - \gamma I \\ \frac{dR}{dt} &= \gamma I \end{aligned} \tag{2}$$

where the disease transmission rate $\beta > 0$ 0 and the recovery rate $\gamma >= 0$ (or in other words, the duration of infection D = 1/ν).

The SIR model, Equation 2, shown in Figure 2, and all variations describe the spread of a virus as a set of first-order differential equations. However, analyzing time responses of output functions, we can assume that we will obtain a better mathematical model of the spread by using a higher-order differential equation to describe the dynamic behavior of the spread.

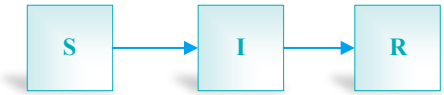

Figure 2. SIR model

The system identification approach is quite different, and the main reason for this is that in all variations of SIR models, differential equations are first-order differential equations, while control systems theory-based models describing the spread of SARS-CoV-2 can be differential equations of higher order, which may be expected to provide a better mathematical description. In addition, advantages of the control system theory approach can also be found in the detailed mathematical analysis of an obtained model regarding observability, controllability and predictions.

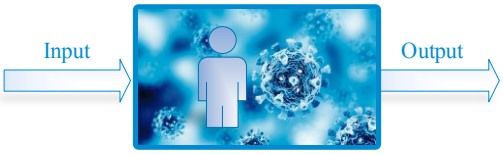

Figure 3. Illustrated black-box model.

A general system identification algorithm and cycle was modified. In this consideration, available official public data were used. We then applied the following steps in control system theory: data pre-processing, fitting the data model, model validation and model analysis. Models that consist of a limited number of inputs and outputs are called black-box models, Brooker et al. (2012). An illustrated black-box model is shown in Figure 3. Note that a problem of system identification is when input and output variables are known, and the system is unknown. The three fundamental problems in physiological control system analysis are prediction, diagnosis, and identification, Khoo et al. (2018). The focus of interest for black-box models is on their input and output characteristics, Ljung (2002).

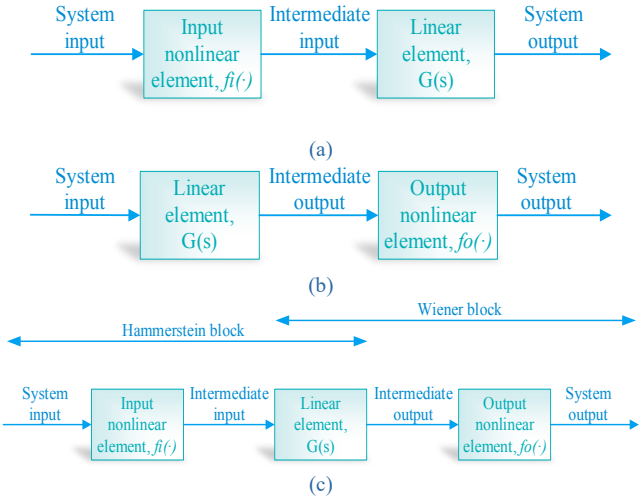

Figure 4. Hammerstein model (a); Wiener model (b); Hammerstein - Wiener model (c).

The spread model was estimated as a transfer function, state-space, process, and nonlinear model. After detailed analysis, we adopted a nonlinear model, namely a nonlinear Hammerstein – Wiener model, for mathematical description of the pandemic spread.

The Hammerstein – Wiener model is block-structured and represents a combination of the Hammerstein model and the Wiener model. The Hammerstein model is based on a static single-valued nonlinear element and a linear dynamic element in series, Giri et al. (2002). The Wiener model is the reverse of the Hammerstein model. A linear element occurs before a static nonlinear characteristic, Ljung (2002).

Thus, a model with a static nonlinearity at the input is a Hammerstein model, and a model with static output is termed a Wiener model. The combination of both, the Hammerstein model and Wiener model is a Hammerstein-Wiener model, Figure 5, Aryani et al. (2014).

The spread model obtained for the SARS-CoV-2 model is developed as such a Hammerstein – Wiener model. The input is the number of tested individuals daily, and the output is the number of infected individuals, Figure 5.

For this purpose, we used official, public data from the European Union Open Data Portal and European Centre for Disease Prevention and Control to define the input and output functions. The website of the European Centre for Disease Prevention and Control (2021) provides various COVID-19 data sets including: vaccination data with data on Covid-19 vaccination in the EU/EEA, daily data with the daily number of new reported COVID-19 cases and deaths by the EU/EEA countries, weekly data on SARS-CoV-1 variants in the EU/EEA, data on 14-day notification rate of new COVID-19 cases and deaths, data on the daily subnational 14-day notification rate of new COVID-19 cases, the weekly subnational 14-day notification rate of new COVID-19 cases, the hospital and ICU admission rates and current occupancy for COVID-19, data on testing for COVID-19 by week and country, the country response measures to COVID-19, the 14-day age-specific notification rate of new COVID-19 cases, as well as data for the maps in support of the Council Recommendation on a coordinated approach to the restriction of free movement in response to the COVID-19 pandemic in the EU/EEA and archived data, including historical data (until December 14th 2020) on the daily number of new reported COVID-19 cases and deaths worldwide.

In this work, we have chosen Serbia as a sample country. Let us assume that number of tests made is an input function of a system, and the number of new cases per day is an output function. The dataset was available from the first registered case in Serbia, from March 6th, 2020. Those vectors are defined in the time domain, and the sampling time is one day, Figure 5. The final value used for the input and output dataset was the last update on December 7th, 2021.

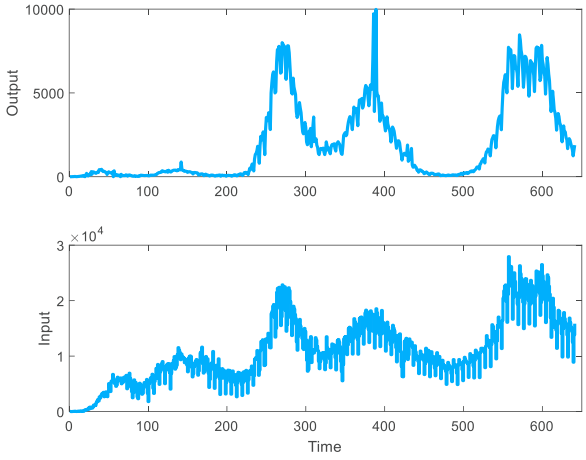

Figure 5. Input vector and output vector, a spread model.

In our consideration, we did not pre-process data before model estimation. As a result, the estimated model compared with measured data is shown in Figure 6.

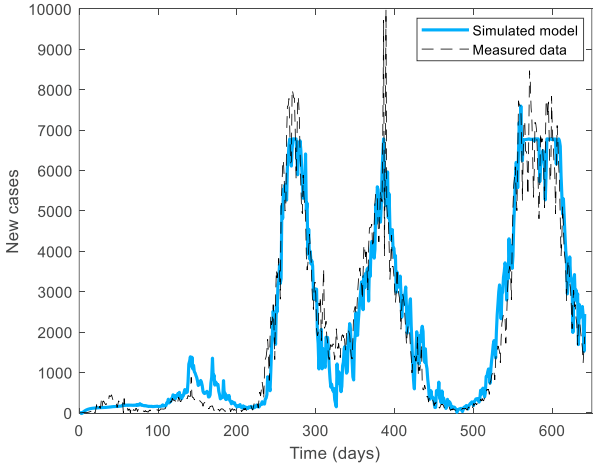

Figure 6. Estimated spread model of SARS-CoV-2 in Serbia.

Since we already defined input and output vectors in MATLAB Workspace, the vectors were loaded in the System Identification App in MATLAB in the time domain. Loaded data can be pre-processed, especially if the disturbance of the data is noticeable. Finally, the estimated model configuration is specified as a single input single output model with input nonlinearity, piecewise linear with ten breakpoints and output nonlinearity, piecewise linear with ten breakpoints, and linear block has two zeros three poles. The results show that the model fits the operational data up to 87% if we ignore the error evident in the real-time output data. The simulated model parameters are given in Table 1, Table 2 and Eq. (3).

Input nonlinearity is defined with ten breakpoints, Table 1,

Table 1. Input nonlinearity breakpoints

Breakpoints	x	y
1.	-2498.8	13487.2
2.	-1028.3	350.8
3.	6118.1	491.1
4.	9858.0	1374.9
5.	13339.4	6103.9
6.	16699.9	5374.6
7.	17577.9	12368.9
8.	18234.2	7721.9
9.	24653.0	11799.3
10.	26608.7	21556.6

The linear block is defined as a discrete-time OE model,

$$y(t) = [B(z)/F(z)\,]u(t) + e(t)$$

$$B(z) = z\text{^}(-1) - 0.9754z\text{^}(-2)$$

$$F(z) = 1 - 1.151z\text{^}(-1) - 0.3892z\text{^}(-2) + 0.5493z\text{^}(-3) \quad (3)$$

Output nonlinearity is defined with ten breakpoints, Table 2,

Table 2. Output nonlinearity breakpoints

Breakpoints	x	y
1.	-17007.5	162963.9
2.	-4789.7	64840.7
3.	1.7	-64.8
4.	2200.7	351.4
5.	2202.0	333.3
6.	28355.2	6781.2
7.	42035.6	6760.2
8.	42814.0	-58946.4
9.	43307.1	-89261.7
10.	46205.3	5702.9

Furthermore, the transient response and pole-zero map of the model obtained show that the model is stable. Following this conclusion, the considerations of observability and controllability considerations can be made. In previous months, the pandemic model was stabilized by imposing restrictions of various constraints, and we can interpret that pandemic can be controlled in an open-loop with proportion action, Figure 7.

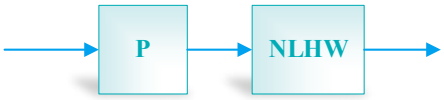

Figure 7. Open-loop control of the spread SARS-CoV-2.

This method applies to data sets in countries where we can assume that the number of tested individuals is the number of infected or susceptible individuals and represents the input function. However, many countries have started mass testing in 2021, and therefore data-driven models become more complicated if incomplete datasets are analyzed.

We have also tried to estimate models for other countries, but it is no longer possible to obtain the model in those countries that have introduced massive testing and are working according to the available data sets. This is because the models are highly data-driven and data-dependent. Estimating a model is very difficult when the data are not consistent, mainly because this means that we have only used a subset of the key elements in the algorithm for system identification.

The system identification algorithm can be used to obtain and develop a mathematical model that correlates the number of deaths and new confirmed cases of infected persons with the novel coronavirus. The input vector is new confirmed cases of infected persons daily, and the output vector is the number of deaths daily, Figure 8.

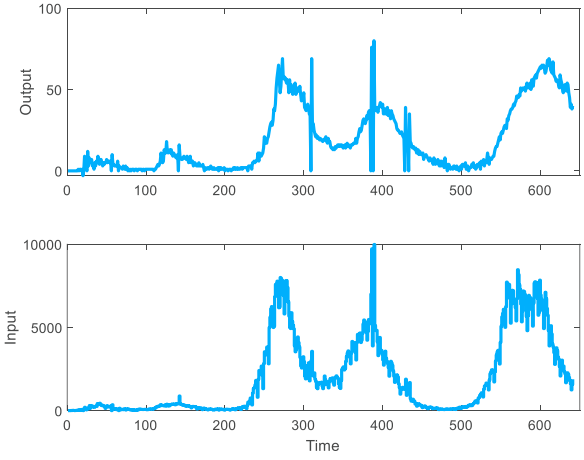

Figure 8. Input vector and output vector, mortality.

We used available data from the Institute for Public Health – Reports, Coronavirus Serbia and Our World in Data, Statistics, and Research, Coronavirus data on the daily notification rate of new COVID-19 cases and deaths to identify the mathematical relationship. For this analysis, the daily number for Serbia was available from March 6th, 2020, to December 15th, 2021. Several models were considered, but after detailed analysis, we decided to use the linear third-order time-invariant model. The model is described with a transfer function, shown in Figure 9. The transfer function G(s) by definition is the ratio of the Laplace transformed output to the Laplace transformed input when all initial conditions are zero, Ogata (2009).

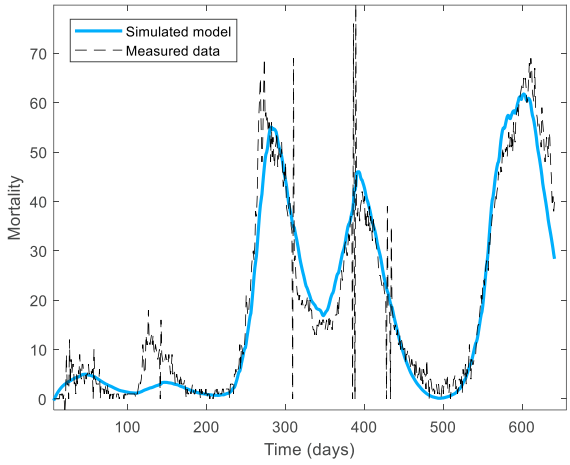

Figure 9. Transfer function model - new confirmed cases vs the number of deaths.

The model after estimation is given by the transfer function in the continuous-time domain,

$$G(s) = \frac{0.0005097\ \mathrm{s} + 1.573\mathrm{e}-05}{s^2 + 0.07792\ \mathrm{s} + 0.001727} \quad (4)$$

The estimated transfer function G(s) model has two poles and one zero at zero initial conditions, and fits the data up to 86% for the data for Serbia. The model developed was also tested for additional countries, the percentage of fitting data, however, was not satisfactory. As expected, this initial model cannot be generalized due to regional differences in the spread of the pandemic and needs to be adapted for each country or region.

With these input and output vectors defined, a correlation was made with the classical SIR model by determining the dynamics of pandemic spread and the number of recovered or deceased people, Figure 10. Correlation is based on results given in Table 1, Table 2, Equation 3, and Equation 4.

Figure 10. Correlation model to SIR model.

An important feature of the model developed for estimating mortality based on the number of confirmed cases is that it is linear and described by the transfer function, and it corresponds to a second-order differential equation. The model enforced stability, and controllability and observability of the model can be considered for further analysis.

6. CONCLUSIONS

The system identification methodology can be used to obtain and develop mathematical models in biomedical sciences. We presented the possibilities of applying methods from systems and control theory for epidemiological questions. The developed models provide a detailed mathematical description of the dynamic behavior since the differential equations are of second, third, or higher than in classic epidemiological models with first-order differential equations. Estimated models are digital and prepared to further analysis and simulation in MATLAB and Simulink.

Data availability is the main problem in the control system theory approach and system identification. The widely used models are data-driven, and if the data is not consistent, it is almost impossible to obtain a reliable model.

The model of the spread of the SARS-CoV-2 was determined as a black-box model which was classified as the Hammerstein –Wiener model. The major drawback of black-box modeling is that the parameters of these estimated models have no physical meaning in terms of equivalence to the process parameters, Zhang P. (2010). Nevertheless, other system characteristics can be observed, such as overshoot, peak response, steady-state, rise time and settling time.

The novel coronavirus mortality model is lower order than the spread model, and is described as a linear model with transfer function and ordinary differential equation.

Note that the estimated models do not include vaccination data, but in future research, depending on available data, these types of data could also be included in the model estimation and validation.

The estimated models developed in this work can be used to predict or develop control strategies of the pandemic and simulate different spread scenarios.

REFERENCES

WHO, World Health Organisation, www.who.int/news-room/q-a-detail/q-acoronaviruses, (accessed on 16.12. 2021)

Ljung L. (2002). *System Identification - Theory for the User*, Second Edition, Prentice-Hall International, ISBN 0-13-656695-2

Dorf R., Bishop R. (2010). *Modern Control Systems*, 12th edition, Prentice-Hall, ISBN: 978-0136024583

Schoukens J., Pintelon R., Rolain Y. (2012). *Mastering System Identification in 100 Exercises*, Wiley-IEEE Press.

Ljung L. (2014). *System Identification Toolbox™ User's Guide*, The MathWorks, Inc.

Verhaegen M., Verdult V. (2007). *Filtering and System Identification, A Least Squares Approach*, Cambridge University Press.

Ogata, K. (2009). Modern Control Engineering - 5th edition, ISBN 978-0136156734

Brooker G. (2012). Introduction to Biomechatronics, University of Sydney, Scitech publishing, inc. ISBN: 978-1891121272

Khoo M. (2018). *Physiological Control Systems – Analysis, Simulation and Estimation*, Second edition, IEEE Press

Series in Biomedical Engineering, IEEE Press, Wiley. ISBN: 978-1119055334

Giordano, G., Blanchini, F., Bruno, R. (2020). *Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy*. Nat Med 26, 855–860, 2020, doi.org/10.1038/s41591-020-0883-7

Weiss H.H. (2013). The SIR model and the foundations of public health. Mater. Mat. 2013, 1–17.

Aryani D., Wang L., Patikirikorala T. (2014), Control Oriented System Identification for Performance Management in Virtualized Software System, IFAC World Congress, 4122-4127, DOI:10.3182/20140824-6-ZA-1003.01100

EU Open Data Portal, EUODP, https://data.europa.eu, (accessed on 03.09.2021)

European Centre for Disease Prevention and Control, https://www.ecdc.europa.eu/en/covid-19/data, (accessed on 16.12.2021)

Giri F., Chaoui F.Z., Haloua M., Rochdi Y., Naitali A. (2002). Hammerstein model identification, Proceedings of the 10th Mediterranean Conference on Control and Automation - MED2002, Lisbon, Portugal, July 9-12.

Institute for Public Health – Reports, Coronavirus Serbia, covid19.rs/ (accessed on 16.12.2021)

COVID-19 statistics in Serbia, available at: covid19.data.gov.rs (accessed on 17.12.2021)

Covid-19 Data Science, available at: https://www.covid-datascience.com (accessed on 17.12.2021)

Our World in Data, Statistics, and Research, Coronavirus Pandemic (COVID19), ourworldindata.org/coronavirus (accessed on 16.12.2021)

Lozanovic Sajic, J., Langthaler, S., Stoppacher, S., and Baumgartner, C. (2021). Analysis of a mathematical model of the spread of the SARS-CoV-2 pandemic determined by the transfer matrix. In G. Müller-Putz, & C. Baumgartner (Eds.), *Proceedings Annual Meeting of the Austrian Society for Biomedical Engineering 2021: ÖGBMT 2021* (pp. 91-94). Verlag der Technischen Universität Graz, doi: 10.3217/978-3-85125-826-4-24

Lozanovic Sajic J. and Djurovic-Petrovic M. (2021). Black-box modeling the spread of Covid-19 in Serbia. In Proceedings of the 3rd Virtual International Conference Path to a Knowledge Society-Managing Risks and Innovation Mathematical Institute of the Serbian Academy of Sciences and Arts.

Zhang P. (2010). Advanced Industrial Control Technology, Chapter 19 - Industrial control system simulation routines, William Andrew Publishing, 781-810, ISBN 9781437778076

Smith D., Moore L. (2021). The SIR Model for Spread of Disease - The Differential Equation Model. Available at: www.maa.org/press/periodicals/loci/joma/the-sir-model-for-spread-of-disease-the-differential-equation-model